

www.qconferences.com
www.qconbeijing.com
www.qconshanghai.com

QCon

伦敦 | 北京 | 东京 | 纽约 | 圣保罗 | 上海 | 旧金山

London · Beijing · Tokyo · New York · Sao Paulo · Shanghai · San Francisco

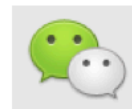
QCon全球软件开发大会

International Software Development Conference

InfoQ^{ueue}



@InfoQ



infoqchina

软件
正在改变世界!

www.qconferences.com

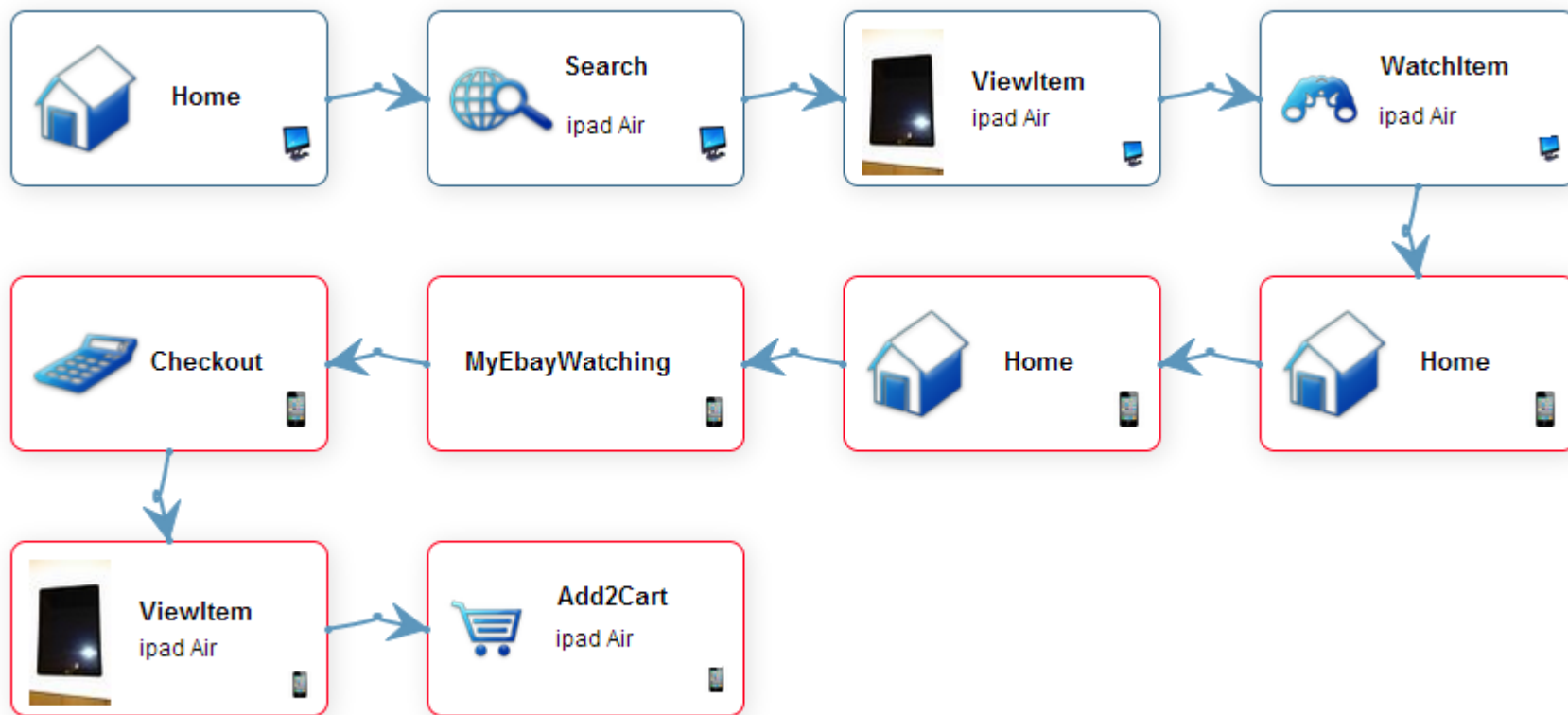


eBay用户行为数据流 实时处理系统

eBay CCOE

汪兴朗

用户行为数据流介绍



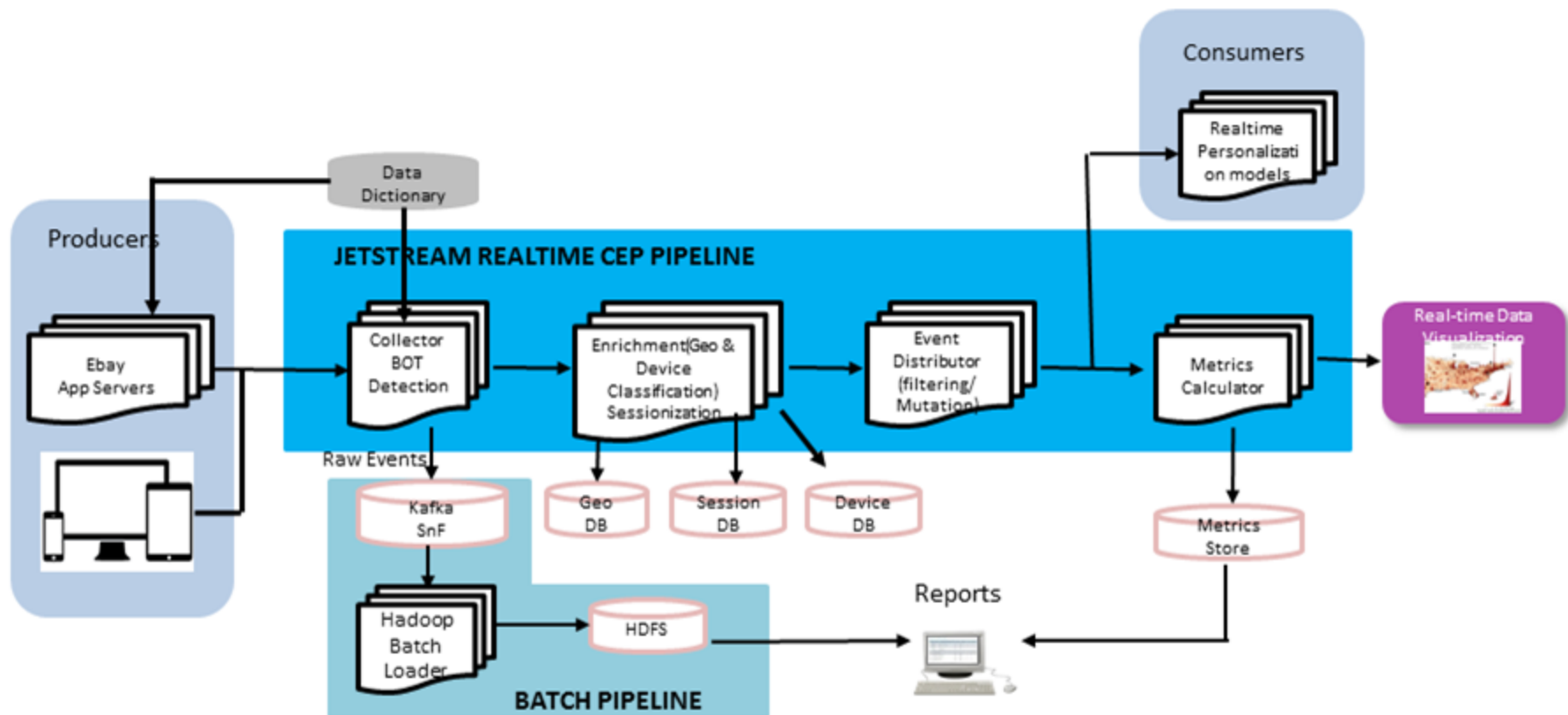
背景介绍

- eBay作为全球性的商务平台和支付行业领先者,拥有海量的用户行为数据
- 用户行为数据拥有巨大的价值
- eBay对用户行为数据处理经历过从DW到 Hadoop的转变
- 现有的Hadoop不能支撑实时性高的业务需求

eBay 用户行为数据特点

- 海量的持续的数据流
- Bot和真实用户数据混合在一起
- 基于用户会话的数据分析
- 移动和桌面端的用户会话的合并
- 7x24 监控和运维

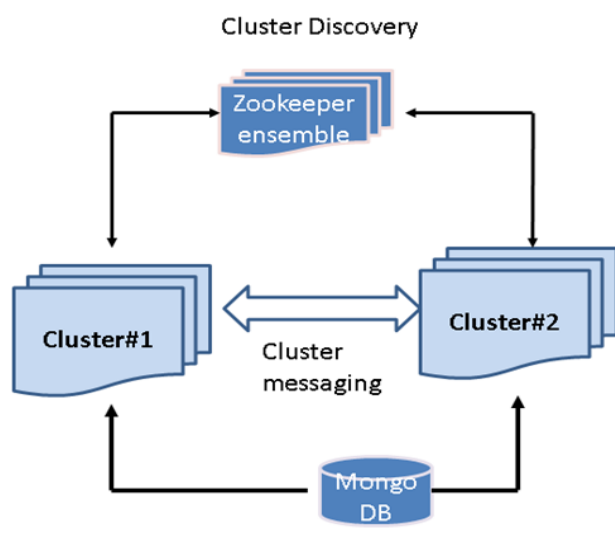
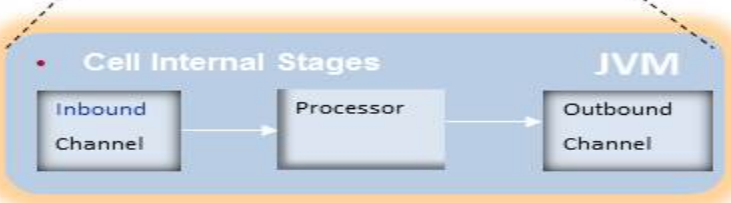
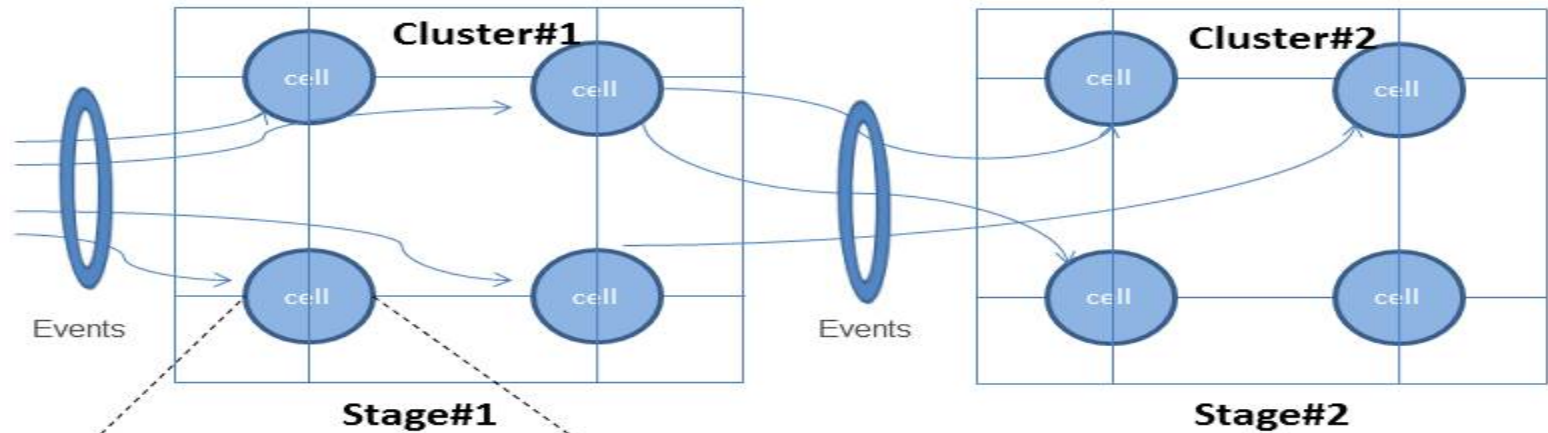
整体架构



关键技术

- Jetstream
 - Pipelining
 - Stream processing & CEP engine (Esper)
 - Broker less messaging
 - Clustering
- Offheap cache and Queue
 - Session state management
 - Store and forward
- Kafka
 - Store and forward, bridge to hadoop
 - Guaranteed messaging with replay capabilities

Jetstream



Jetstream

- Spring as container
- Zookeeper for coordinator
- Netty4 for async messaging
 - Batching
 - Compress
- Kryo as serialization
- Disruptor as Queue
- MongoDB for configuration
- Affinity based messaging routing
- Esper for declarative rules

Why Jetstream?

- Open/Auto-discover topology
- Endless stream
- Push model
 - Measure loss rather than blocking
 - Proactive throttling
- Declarative event processing logic
- Cost effective messaging
- Predictable latency

Offheap 技术的应用

- 应用场景
 - Sessionization
 - In-memory data aggregation
 - Bot detection
- OffHeap Cache
 - Low latency
 - Concurrent by partition cache by key
 - $O(1)$ read /write by hash
 - Instantaneous expiration by an ordered linked list, no full scan required, $O(1)$ for session expiration
 - Minimized memory footprint by optimizing serializer
 - TimeSlot to support diverse session TTLs

Kafka的运用

- 应用场景
 - Online stream to offline Hadoop
 - Temp store for guarantee messaging processing
- Kafka Replay
 - Upstream write message to kafka when it can not push to downstream.

运营经验

- Rest API for metrics
- SiteBot for alerting by check metrics
- TSDB for metrics data
- MongoDB for configuration change

应用案例

- In-session personalization
- Site monitoring
- Advertisement
- Experimentation
- Attribution analysis
- Real-time reporting

Real-Time Web Checkouts

Checkout Item Samples

DENVER, USA
HOUSTON, USA
LEATHERHEAD, GBR
SAN MATEO, USA
CORIO, AUS



 SHARE

Courtesy of PubNub, [NASA](#), [JHT](#), and [Spacescape](#)

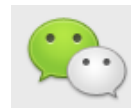




Brought by **InfoQ**[®]



@InfoQ



infoqchina

软件
正在改变世界!